

Dear Sprocket Central,

After looking over the datasets you sent, we found a few significant quality issues that will affect our data analysis. The problems with the data quality and the corrective measures taken are described below. To increase the accuracy of your data and enable you to make wiser business decisions, consider the following recommendations-

List of Issues

(1) Accuracy :

Issues: A few dates of birth in the "Customer Demographic" section are not correct, and the "Transactions" sheet lacks a profit column.

Recommendation: To ensure correctness, create a "profit column" in Transactions. Future analyses will also benefit from it.

Actions Taken: Filter out the outlier DOB.

(2) Completeness :

Issue 1: There were discrepancies in the additional customer IDs for "Customer Address," "Customer Demographic," and "Transactions."

Recommendation: Check that the tables are up to date. Only IDs ranging from 1 to 3500 will be utilised since they include complete data.

Actions Taken: Filter all customer ids from 1 to 3500.

Issue 2: Blanks in the online order and brand column for "Transactions," as well as in the job title for "Customer Demographic."

Mitigation: Filter out 'blanks' for job title, online order, and brand column.

Recommendations: Modify the job title to a different category, like industry, or give the job title dropdown options. Give the brand column and online order dropdown options. Blanks are incomplete data that can distort the outcomes of additional analysis. Dropdown options will make it possible to have more complete data, which will lead to more accurate analysis.

Actions Taken: Filter out 'blanks' for job title, online order, and brand column.

(3) Consistency :

Issues: gender is inconsistent with "Customer Address" and "Customer Demographic," respectively.

Actions Taken: Change every "M" to "male." For gender, change every "F" and "Female" to "Female." For states, filter all of Victoria to "VIC" and all of New South Wales to "NSW."

Recommendation: Make choice menus with the gender options "Male," "Female," and "U." Make a dropdown menu with all of the state abbreviations.

(4) Currency:

Issues: Individuals with a deceased indicator of 'Y' are not currently customers according to "Customer Demographics."

Actions Taken: Remove clients who have selected "Y" in the deceased indicator.

Recommendations: Although it can be challenging to search for deceased customers, data should be updated as soon as this information is obtained. Since deceased customers are no longer active, their removal from the data will make it more current and produce more precise estimates for analysis in the future.

(5) Relevancy:

Issues: The order status for "Transactions" and the default column for "Customer Demographic" are not clear or relevant.

Actions Taken: Metadata derived and placed in default columns. Remove the "Cancelled order status" filter.

Recommendation: Look for anything unclear. To make the data comprehensible, remove or format the metadata. For upcoming analyses, the status of cancelled orders is irrelevant because it can skew the data. For instance, the annual total number of customers will be overstated.

(6) Validity:

Issues: List price format and product sale date format for "Transactions" sheet.

Actions Taken: Format the list price to the appropriate currency and the sale date to a short date format.

Recommendations: When adding new data, set up columns so that formats like price and decimals are already there. Allowable values will facilitate easier interpretation of the data. Readability will be improved by formatting into the price and allowing for two or three decimals to be placed consistently. This will have a positive impact on the accuracy and speed of analysis used to make business decisions.

All the data quality issues identified in the first step of data quality analysis is summarized here. The suggested actions and strategies are simple and effective ways to improve data quality for future analysis. Not only do they improve the results of analyzes that can be performed in-house, but they also increase the scope of analyzes that can be performed by KPMG and other contracted analysis teams.

Thank you.